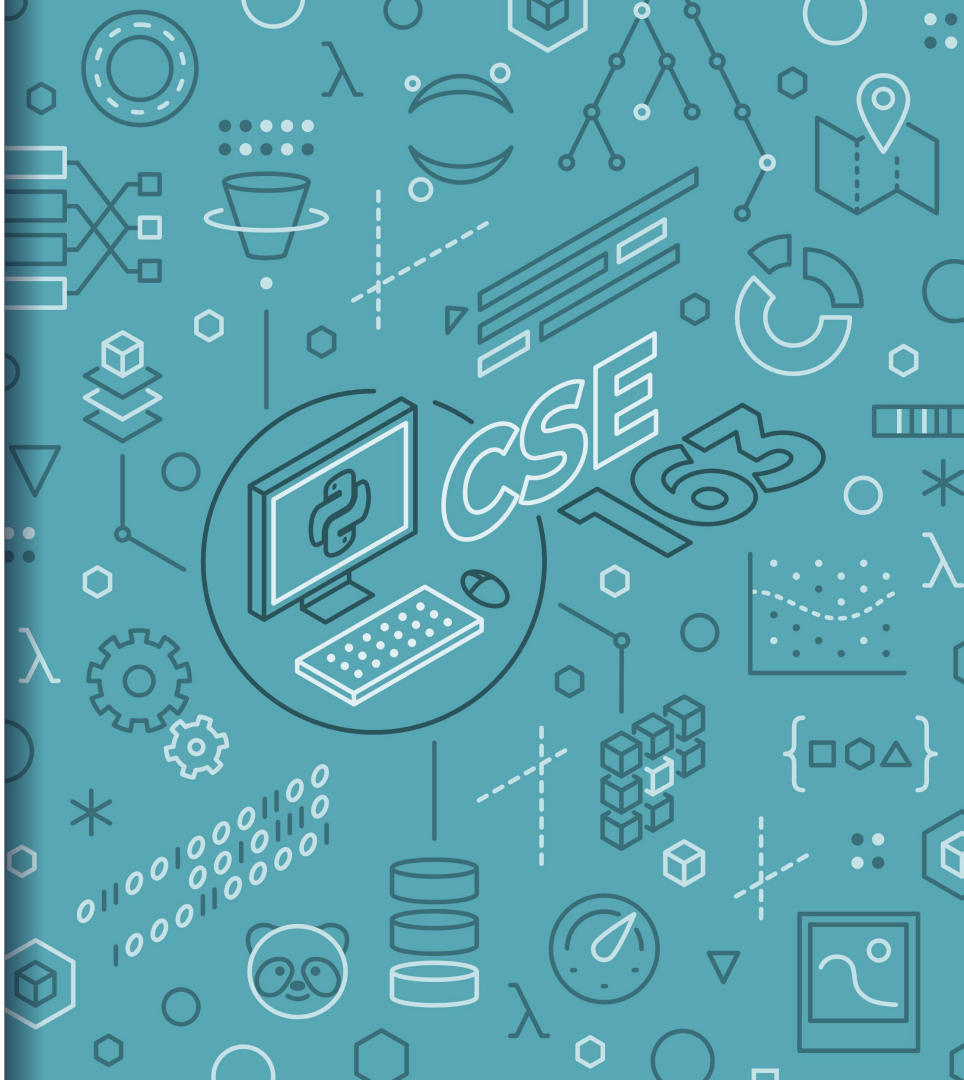
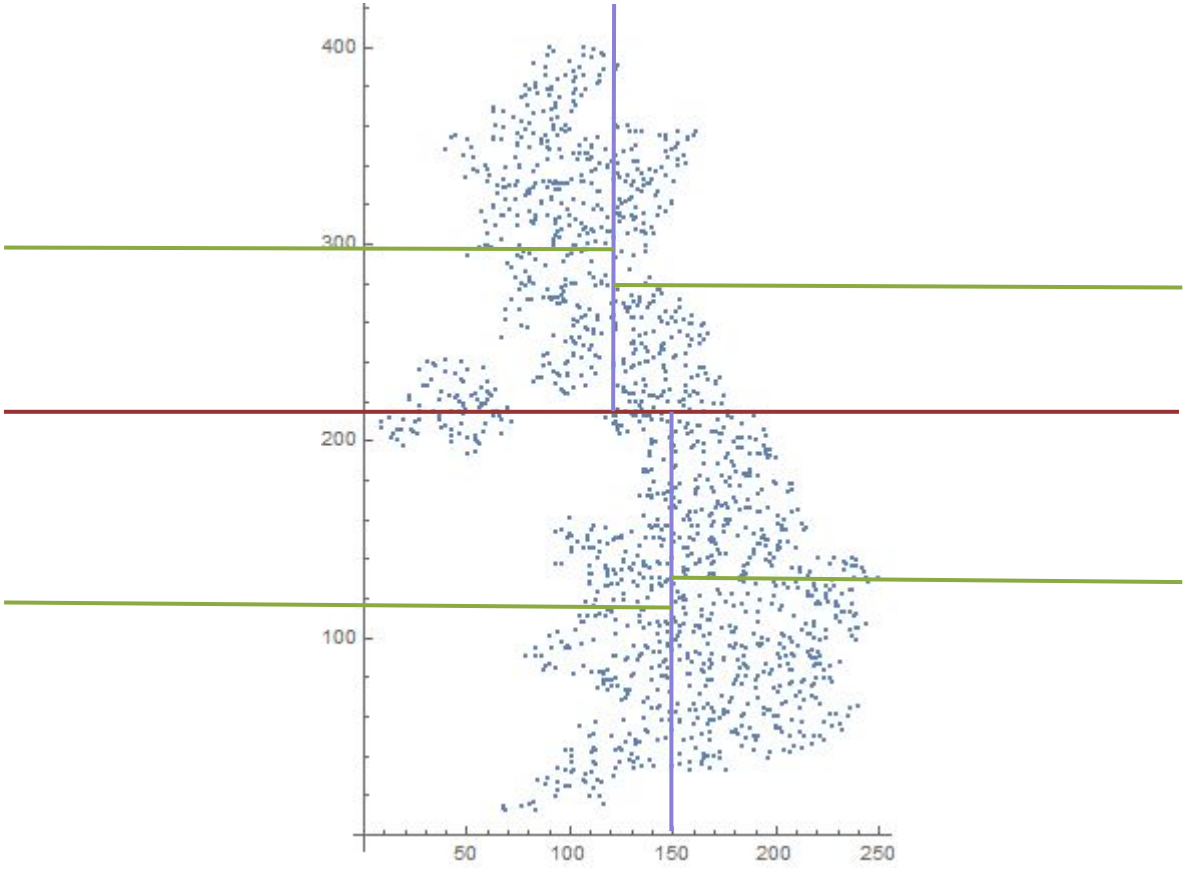
CSE 163

Ethics

Hunter Schafer

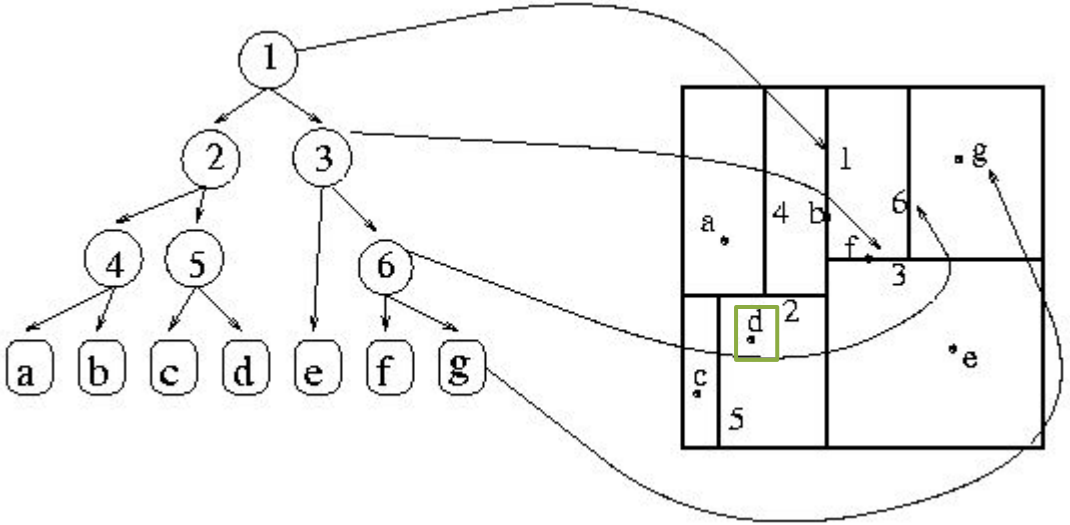


Spatial Index



Spatial Index

This is a tree!



To find the points in a region, just follow the tree

Ground Rules

- Talking about ethics and data science can be kind of hard since it can impact our lives in so many ways.
- I have some case studies I want to talk us through
  - Can give you reference points to think about on whatever project you are doing.
- With a field this new, there aren't a lot of "right answers"!
  - It's just a lot of people thinking hard about this stuff and discussing what they think. We will do that today! :)

- Do remember that I am just one voice in the room and other people likely have different opinions than me (or you!).
- We should remember to be respectful in our discussions and make sure we are asking critical questions of what we believe.

Case Study 1

Uber

Rides of Glory

- 2012 blog post describing how they could predict “hook ups” based on ride data.
- Data presented was in aggregate (i.e. no one was called out)
- Is the unethical?

Data Usage

- Users give their data to applications (willingly?)
- It's our responsibility to make sure we **respect** that data
- Ask yourself these questions:
  - What consent was given when the data was collected?
  - Would the user want that data private? Would they be happy to know you used it in that way?

- Data Science Principles
  - Keep user data anonymized and secure
  - Know where the data came from and what permission we have to use it

- Further resources
  - k-anonymity / Differential privacy
  - Security

Case Study 2

COMPAS

- A machine learning system made by Northpointe to help predict if someone leaving jail will commit a crime again
  - This is called recidivism
- ProPublica did an analysis of COMPAS results and came to the conclusion COMPAS was racist!
  - They found that it was much more likely to predict high recidivism for black people than white people
- Northpointe countered that their scores are accurate, this agreed with ProPublica's finding that if someone was given a high score, they had mostly the same chance of recidivism despite the race of the person
- Is it unethical to use this model? What makes a model fair?

Fairness in Machine Learning

Many definitions of fairness

- Most contradictory!

	Definition	Paper	Citation #	Result
3.1.1	Group fairness or statistical parity	[12]	208	×
3.1.2	Conditional statistical parity	[11]	29	✓
3.2.1	Predictive parity	[10]	57	✓
3.2.2	False positive error rate balance	[10]	57	×
3.2.3	False negative error rate balance	[10]	57	✓
3.2.4	Equalised odds	[14]	106	×
3.2.5	Conditional use accuracy equality	[8]	18	×
3.2.6	Overall accuracy equality	[8]	18	✓
3.2.7	Treatment equality	[8]	18	×
3.3.1	Test-fairness or calibration	[10]	57	✓
3.3.2	Well calibration	[16]	81	✓
3.3.3	Balance for positive class	[16]	81	✓
3.3.4	Balance for negative class	[16]	81	×
4.1	Causal discrimination	[13]	1	×
4.2	Fairness through unawareness	[17]	14	✓
4.3	Fairness through awareness	[12]	208	×
5.1	Counterfactual fairness	[17]	14	–
5.2	No unresolved discrimination	[15]	14	–
5.3	No proxy discrimination	[15]	14	–
5.4	Fair inference	[19]	6	–

Table 1: Considered Definitions of Fairness

Verma and Rubin, FairWare2018

Fairness with COMPAS

A simpler version would be three ideal things for COMPAS

- Equally accurate in predicting recidivism across sub-groups
- Assume members for different groups have the same chance of being wrongfully predicted to recidivate
- Failure to predict recidivism happens at the same rate across groups

People have shown that it's not possible to satisfy all 3 in most real world scenarios (unless the groups are exactly the same).

Algorithm Bias

- So is the COMPAS algorithm racist?
  - Well, no... It's not a person and has no sentience.
  - But also, yes... It impacts different races differently.
  - It's not the algorithms fault! It's the data's fault.

```
text_to_sentiment("My name is Emily")
# 2.2286179364745311
text_to_sentiment("My name is Heather")
# 1.3976291151079159
text_to_sentiment("My name is Yvette")
# 0.98463802132985556
text_to_sentiment("My name is Shaniqua")
# -0.47048131775890656
```

Fixing Bias

- How can we fix this?
  - No “race” column?
    - Many things are correlated with race.
  - Add a “race” column and try to protect it with a notion of fairness?
    - Illegal in many cases!
- Explainable Models: Use models that can “explain” why they made the choices they did (Decision Tree!)
- If possible, find a more representative or fairer dataset

Further Reading

- A new field of fairness and ML (FAT* Conference)
- The Ethical Algorithm - Michael Kearns and Aaron Roth

Case Study 3

Baltimore Potholes

- The city of Baltimore has a problem with potholes
- People were able to report potholes to the city, but it was a slow process
- Instead, they decided to make an app that will detect if someone drove over a pothole using the phone sensors
- Can use some data science techniques we learned in this class to try to predict where potholes are.
- Is this an ethical use of data science?

Access

- This sounds great in theory, but think about the population of people that have smartphones
  - More biased towards affluent neighborhoods
  - Fear of “leaving behind” communities that don’t have the same access to technology
- An unreasonable assumption is to assume all people have the same amount of access to the same opportunities
  - Race, gender, sex, disability, etc.
  - Where people live
  - What technologies they can afford

Key Takeaways

- Try to pay attention to the world around you and see if people are using data science ethically
  - If they aren't ask what assumptions did they make that lead them to the wrong conclusions?
  - It's rare that people are acting maliciously, but rather didn't think about all the cases
- Ask yourself hard questions about your projects
  - Who are the stakeholders? Who will this help? Who might the potentially hurt?
- Whenever you are doing statistics or modelling, there will always be uncertainty and errors. Think about what errors are tolerable and which aren't.

Uncertainty in Models

“All models are wrong, but some are useful” - George Box

In your application, surface the idea of uncertainty to the user